DOCUMENT RESUME

ED 251 486 TM 840 762

AUTHOR	Frankel, Steven M.; And Others
TITLE	Study of Test Burden at the Elementary and Intermediate Schools.
INSTITUTION	Montgomery County Public Schools, Rockville, Md. Dept. of Educational Accountability.
PUB DATE	May 83
NOTE	21p.; Small print in some tables.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Achievement Tests; *Educational Testing; Elementary Education; Group Testing; Junior High Schools; School Districts; *Standardized Tests; Test Coaching; Test Reliability; Test Use; Test Validity; *Time Management
IDENTIFIERS	*Montgomery County Public Schools MD

ABSTRACT

In response to concerns raised by principals at both the elementary and intermediate levels regarding the amount of instructional time consumed by students taking tests which were not related to day-to-day instruction, a study was conducted of the amount of time spent preparing to give such tests and actually administering them, and the manner in which the test results were being used. The average number of hours spent preparing for and taking these tests ranged from 6.0 hours in Grade 6 to 29.5 hours in Grade 3. These results indicate that the amount of periodic testing is not large and in fact consumes less than 2 percent of the available instruction time. Nonetheless, three areas of concern emerged from the analysis: (1) the status of locally-developed curriculum-related tests, given teacher uncertainty about their usefulness and quality; (2) the misuse of achievement test data by using results to place individual students in particular programs and classes; and (3) questions about the reliability and validity of the measures used, especially at the intermediate level. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED251486

MONTGOMERY COUNTY PUBLIC SCHOOLS
Rockville, Maryland

STUDY OF TEST BURDEN AT THE ELEMENTARY AND INTERMEDIATE SCHOOLS

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Hymes

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 762

May 1983

Edward Andrews
Superintendent of Schools

MONTGOMERY COUNTY PUBLIC SCHOOLS
Rockville, Maryland

STUDY OF TEST BURDEN AT THE ELEMENTARY AND INTERMEDIATE SCHOOLS

By
Steven M. Frankel
Joy A. Frechtling
N. James Myerberg
Pam Splaine

Steven M. Frankel, Director
Department of Educational Accountability

Joy A. Frechtling, Director
Division of Instructional
Evaluation and Testing

STUDY OF TEST BURDEN AT THE ELEMENTARY AND INTERMEDIATE SCHOOLS

In response to concerns raised by principals at both the elementary and intermediate levels regarding the amount of instructional time consumed by students taking tests which are not related to day-to-day instruction, the Department of Educational Accountability (DEA) conducted a brief study of:

1. The amount of time spent preparing to give such tests
2. The amount of time spent actually administering them
3. The manner in which the test results are being used

Using a sample of 12 elementary and 6 intermediate schools, about 115 school staff members were interviewed regarding these issues. In elementary schools, principals provided the most information; but additional data were obtained by interviewing an early primary, third grade, fifth grade, resource room, and reading teacher in each school. In intermediate schools, information was obtained from the principal, a counselor, and an English, math, science, social studies, and physical education teacher. Where possible, resource teachers responsible for overall department operation in these subject areas were interviewed. Selected area and central office staff were also interviewed.

For the purposes of this study, only data relating to tests used to assess a broad range of skills on a periodic basis were considered. Included in this category are the tests mandated by the State Department of Education; some additional standardized tests mandated by MCPS; some tests administered in the fifth, sixth, and eighth grades to facilitate the placement of students in junior and senior high courses; criterion-referenced measures administered as part of the prototype Instructional Program in Reading/Language Arts (IPR/LA); science and social studies instructional programs; and a variety of tests used in individual schools to assess student progress in specific subject areas once or twice during the school year.

Not included within the scope of this study were tests used with individual students to assess special needs and tests which are used on an ongoing basis as an integral part of the regular instructional program. These exclusions include tests designed by classroom teachers to assess progress on a weekly basis (e.g., Friday spelling tests, tests on specific chapters or instructional units, etc.), tests developed by publishers or within MCPS to assess progress on a regular basis (e.g., tests included in basal reading series, the Instructional System in Mathematics tests, etc.), and midterms and finals.

One reason for excluding the latter types of tests from the study was that the highly individual nature of their administration, depending on individual student progress or the progress of a group of students, makes deriving general estimates regarding the time demands very difficult. For example, preliminary questions regarding testing time for the ISM indicated that time for testing varied widely for individual students, making it extremely difficult to develop even a grade-level estimate. Factors affecting these time estimates included the rates of progress and number of objectives attempted by individual students, the type of testing used (computer vs. aides), the presence or absence of scheduling problems, and the time taken by travel between the classroom and testing site. Similarly, the time allocated by individual teachers for classroom tests has even more variance. It was,

therefore, felt that the inclusion of such tests would have greatly increased the study's time requirements, complexity, and cost; making it extremely difficult, if not impossible, to obtain a preliminary look at testing burden in a timely manner. However, if more time and dollars are available at a later date, a study of classroom testing, intended for what might be called "ongoing instructional evaluation," will be undertaken, since such an effort is likely to be both interesting and fruitful.

Originally, tests used to select students for gifted and talented programs were also included. However, the reports given by schools differed so widely, especially with regard to time for the Renzulli (from 1 to 55 hours at a single grade level), that it was impossible to develop a coherent picture. This suggests that staff were either unable to recall accurately how much time this screening consumes or that practices were very different in each of the schools sampled.

ELEMENTARY SCHOOL FINDINGS

The major finding of this study, which relates to elementary schools, is that not very much student time is being spent taking the kind of examinations which are described in this study. As detailed in Exhibit 1, the average number of hours spent taking these tests were as follows:

Grade 1	5.5 hours
Grade 2	5.5 hours
Grade 3	14.0 hours
Grade 4	5.5 hours
Grade 5	11.5 hours
Grade 6	5.0 hours

The bumps in Grades 3 and 5 are caused by the California Achievement Tests (CAT), which are state mandated. The California Tests account for approximately 50 percent of the time devoted to testing in these grades. However, assuming that there are about 714 hours available for instruction in a typical school year (178.5 days @ 4 hours per day), even the highest allocation of 14 hours represents only about 2 percent of the available instructional time. If anything, these totals appear low and suggest that the system might well consider adding summative evaluation measures in Grades 1, 2, 4, and 6 so that progress could be assessed in each school annually.

To further analyze these data, a major distinction must be made between time spent actually administering a test and time spent preparing students to perform well on it. In the case of the CATs, where both the school system and individual communities place a great deal of importance on the results, it was not surprising to learn that a significant amount of time was being spent preparing students to take these tests. In fact, as shown in Exhibit 2, it was learned that students in Grades 3 and 5 are spending as much time preparing for the CATs as they spend taking all of these types of tests combined.

According to Exhibit 2, an average of about 13.5 hours was spent in Grades 3 and 5 preparing students to take the California Achievement Tests. This time varies widely across individual schools, however, with preparation time ranging from 1 to 32 hours in the schools sampled.

EXHIBIT 1

Elementary Schools: Time for Testing by Grade Level and Type of Test*

Type of Test	N	Average Administrative Time** (in hours) K	1	2	3	4	5	6	Average Preparation time (in hours) K	1	2	3	4	5	6	Average Total Time*** (in hours) K	1	2	3	4	5	6
		(11)	(11)	(11)	(12)	(12)	(12)	(11)	(11)	(11)	(11)	(12)	(12)	(12)	(11)	(11)	(11)	(11)	(12)	(12)	(12)	(11)
State Required		15.5	1.5	2.0	7.0	1.5	6.0	-	-	-	-	14.0	-	13.0	-	15.5	1.5	2.0	21.0	1.5	19.0	-
MCPS Required		-	-	-	3.0	-	1.5	-	-	-	-	0.5	-	-	-	-	-	-	3.5	-	1.5	-
Receiving School Required		-	-	-	-	-	-	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	1.0
Optional Curric. Related		-	3.5	3.5	3.5	3.5	3.5	3.5	-	1.0	1.0	1.0	1.0	1.0	1.0	-	4.5	4.5	4.5	4.5	4.5	4.5
Optional School Initiated		0.5	0.5	-	0.5	0.5	0.5	0.5	0.5	-	-	-	-	-	-	1.0	0.5	-	0.5	0.5	0.5	0.5
Total		16.0	5.5	5.5	14.0	5.5	11.5	5.0	0.5	1.0	1.0	15.5	1.0	14.0	1.0	16.5	6.5	6.5	29.5	6.5	25.5	6.0

*Some of the times in this table do not match those in Exhibit 2. This is because the number of schools used to compute the averages is different. In this table, the burden is considered for all sample schools with a given grade. In Exhibit 2, the burden is considered only for the schools that actually gave the test.

**Rounded to nearest half hour.

***Includes preparation and administration.

706b/75

EXHIBIT 2

Tests Given in the Elementary Schools*

	No. of Schools**	% of Schools	Adm. Year	Average Administrative Time*** K	1	2	3	4	5	6	Average Preparation Time 1	2	3	4	5	6	Average Time Devoted to Testing**** K	1	2	3	4	5	6
State Required Tests																							
CAT	12	100	1-2		5.5	10.5	7.0	8.5	6.0		0	0	14.0	0	13.0	-		5.5	10.5	21.0	8.5	19.0	
Early Childhood Identif.	12	100	1	15.5	1.0	1.0	1.0				0	0	0	-	-	-	15.5	1.0	1.0	1.0			
MCPS Required Tests																							
AAHPERD	12	100	1						1.5		-	-	-	-	0	-						1.5	
COGAT	12	100	1				3.0				-	-	0.5	-	-	-				3.5			
School Required Tests																							
Holt Reading	1	8	1						2.0	3.5	-	-	-	-	0	0						2.0	3.5
Stanford Achievement	2	17	1							2.0	-	-	-	-	-	0							2.0
Jr. High Math	1	8	1							1.5	-	-	-	-	-	0							1.5
Optional Curriculum Related																							
IPR/LA	10	83	1-2		4.0	4.0	4.0	4.0	4.0	4.0	1.0	1.0	1.0	1.0	1.0	1.0		5.0	5.0	5.0	5.0	5.0	5.0
Map and Globe Skills	1	8	2						2.0		-	-	-	-	0	-						2.0	
Optional School Initiated																							
Boehm	2	17	1	1.0							-	-	-	-	-	-	1.0						
Botel	2	17	1				0.5		0.5		-	-	0	-	0	-				0.5		0.5	
Clymer-Barrett	1	8	1		2.0						0.5	-	-	-	-	-		2.5					
Ginn	1	8	1	0.5							0	0	0	0	0	0	0.5						
Holt Reading	1	8	1				3.0	3.0	3.0	3.0	-	-	0	0	0	0				3.0	3.0	3.0	3.0
Houghton-Mifflin Rdg	2	17	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0	0	0	0	0	0	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Mechanics of English	1	8	1						1.0		-	-	-	-	0	-						1.0	
Metro Readiness	2	17	1	1.5							-	-	-	-	-	-	1.5						
Morrison McCall	2	17	2-3				1.0		1.0		-	-	0	-	0	-				1.0		1.0	
Stanford Achievement	1	8	1		3.0	0.5	0.5	0.5	0.5	0.5	0	0	0	0	0	0		3.0	0.5	0.5	0.5	0.5	0.5
SESAT	2	17	1	1.0							-	-	-	-	-	-	1.0						

*Excludes tests for gifted and talented screening and for assessing individual students with special needs.
**For a given test, the number of schools at different grades may vary.
***Rounded to the nearest half hour.
****Includes both preparation for testing and actual test administration.

706b/75

A possible additional factor is time to score tests. While many of the tests are scored outside the school, others, such as the IPR/LA tests in at leaast some grades, currently require teacher time for scoring and interpreting results. While this activity does not necessarily reduce instructional time, it may well contribute to the overall test burden.

In terms of how much of this testing time could be eliminated at the elementary level, the answer is "very little." The number of tests required by MCPS beyond the state program is minimal and decreasing. The Cognitive Abilities Test, which consumes nearly four hours at the third grade level, will no longer be required after this year. Beyond the tests required by the state, the test requirements are far from demanding.

Turning to the tests that are currently administered at the schools' option, the picture with regard to burden becomes more complex. The amount of optional curriculum-related testing (mostly IPR/LA and Map and Globe Skills) and optional school-initiated testing (mostly reading tests) totals to less than five hours per grade level. However, most of this additional time is devoted to administration of narration tests associated with IPR/LA. While these tests appear sound from several technical aspects (see Appendix A), they are not always used for their intended purpose, in part because many teachers do not fully understand what they are intended to assess. A significant problem here is that very few teachers have received intensive training in these criterion-referenced measures. For instance, data from the Study of Elementary Reading Instruction show that as of the first semester of the 1982-83 school year, only 6 percent of the teachers in our sample schools had received any in-service training on the use of these tests. Further, confusion exists regarding whether the IPR/LA tests are required. In interviews conducted with Central Office personnel, staff was told that administration of the IPR/LA tests is optional. However, memoranda from the Central Office have not always been clear on this matter, and their use has been strongly encouraged, if not required, in two of the three areas.

Thus, many teachers feel they are being required to use tests whose purpose and value are not at all clear to them. The concerns raised by this experience with IPR/LA narration tests have raised additional anxieties regarding other IPR/LA tests and other "optional" curriculum-related tests currently in various stages of development and dissemination. Since only limited information has been gathered about the quality or usefulness of these other measures, it seems advisable that any expansion of them be done very carefully and cautiously, and that decisions as to whether test administration is mandatory or optional be directly communicated to the teachers responsible for teaching the material.

So far, this study has focused mainly on tests administered in Grades 1-6. A special situation exists at the kindergarten level because of the state-mandated Early Childhood Identification (ECI) Program, which requires about half an hour of paperwork by the kindergarten teacher for every child in his/her class. This measure is intended to assist schools in flagging students who may be in need of special services. Several schools reported that they did not use the results of the ECI screening at all; and some teachers felt that the instrument was so poor that, in and of itself, it could not provide useful data for diagnosis of student needs. Since there are no data available to support that this test is either reliable or valid for identification of students at risk, the opinions of these teachers cannot be refuted. Procedures for the Early Identification Program have recently been

revised by the state, and the revised forms will be required to be initiated in all elementary schools next year. Whether these will prove more useful and less burdensome is unknown at present, and it is clear that the time burden they place on teachers must be carefully examined.

In the above discussion, this report has touched briefly on concerns that school staff have raised regarding the use of some of the required and optional tests currently being given. This present study has also reinforced concerns on the part of DEA staff regarding the manner in which the standardized achievement tests included in the elementary school testing program are being used. It was made clear to our interviewers that California Achievement Test (CAT) results and results of other standardized measures are used in some schools as a primary factor in placing students in classes or programs. This is an inappropriate use of these tests since their standard errors of measurement are too large to provide reliable data at the individual student level, and further education of principals and teachers seems to be needed.

Another area of related concern is the use by some elementary schools of standardized achievement tests other than the CAT. In some schools, these are even given in Grades 3 and 5 despite CAT results being available for the same students. While DEA has little faith in the utility of the level of the CAT which the state mandates us to use at the third grade level, the department would still recommend that if additional achievement testing is to be done to assess overall school progress, the CAT should be used. Since the scores made on all levels of the test are linked together statistically, this would provide principals with a better idea of the progress being made in given areas from year-to-year than would using a different test battery altogether. Also, principals might explore using out-of-level testing with the CAT at the grades in which testing is optional. We would repeat, however, that these tests should not be used to assess individual student progress and that measures other than standardized achievement tests should be used for this purpose.

In summary, no major problems are seen requiring concerted action in regard to the amount of periodic testing at the elementary level. While there is clearly room for improvement in some areas, and a close look at the usefulness of the MCPS curriculum-related tests is called for before further expansion takes place, the present testing program consumes only a small fraction of the instructional time available. Most of the activities consist of tests which are mandated by the state and are largely beyond the school system's sphere of control. Finally, much of the additional time associated with testing, such as preparing for the CAT, clearly has some instructional benefits and is under the control of the local school. It must be emphasized, however, that our analysis includes only selected testing and does not address the testing time devoted to ISM assessments, basal tests, and other curriculum-embedded measures used repeatedly throughout the year to assess progress.

INTERMEDIATE SCHOOL FINDINGS

The intermediate schools are impacted more heavily than the elementary schools by periodic testing. Whereas test administration activities consumed from 5 to 14 hours in the elementary grades, Exhibit 3 shows that the total is higher is the intermediate grades, as seen in the following:

EXHIBIT 3

Secondary Schools: Time for Testing by Grade Level and Type of Test*

Type of Test	N	Average Administrative Time** (in hours) 6 (1)	7 (6)	8 (6)	9 (3)	Average Preparation Time (in hours) 6 (1)	7 (6)	8 (6)	9 (3)	Average Total Time*** (in hours) 6 (1)	7 (6)	8 (6)	9 (3)
State Required		-	5.0	6.5	7.0	-	3.5	1.0	3.0	-	8.5	7.5	10.0
MCPS Required		-	-	3.5	1.5	-	-	-	1.0	-	-	3.5	2.5
Receiving School Required		-	-	-	-	-	-	-	-	-	-	-	-
Optional Curriculum Related		5.5	5.5	3.0	-	0.5	1.0	0.5	-	6.0	6.5	3.5	-
Optional School Initiated		-	3.0	3.0	0.5	-	-	-	-	-	3.0	3.0	0.5
Total		5.5	13.5	16.0	9.0	0.5	4.5	1.5	4.0	6.0	18.0	17.5	13.0

*Some of the times in this table do not match those in Exhibit 4 because the number of schools used to compute the average is different. In this table, the burden is considered for all sample schools with a given grade. In Exhibit 4, the burden is considered only for the schools that actually gave the test.

**Rounded to nearest half hour.

***Includes preparation and administration.

708b/75

Grade 7 13.5 hours
Grade 8 16.0 hours
Grade 9 9.0 hours

One reason for the heavier load is the presence of state-mandated tests in each of the intermediate school grades: the CAT in Grade 8 and the Project Basic examinations in Grades 7 and 9. As shown in Exhibit 4, state-required tests consume about 40 percent of these kinds of testing in Grades 7 and 8 and more than 75 percent in Grade 9.

In addition, the intermediate schools in the sample seemed to spend somewhat more time on the optional tests imbedded in the instructional systems (IPR/LA, Map and Globe Skills, and Science Criterion-Referenced Test); and they elect to do more testing using norm-referenced achievement test batteries.

The impact of MCPS-required tests is not large and will decrease next year when the Writing Proficiency Test is eliminated.

An interesting fact which emerges from these data is that much less time is spent preparing students to take the state-mandated tests in the intermediate schools than is the case in the elementary schools. This is understandable in the case of the Maryland Functional Reading Test because so many of our students pass the test the first time they take it. However, this situation will probably change now that the results of the Maryland Functional Mathematics Tests are available, since the first year results show more than half of our seventh graders falling within the range which is considered as being the danger zone for passing the test in the ninth grade. The results also show more than 30 percent of our present ninth graders have failed the test. Given the nature of our school system, this will undoubtedly lead to a reexamination of the mathematics curriculum and mathematics instruction and to much more intensive test preparation activities.

If initial results are poor on the Maryland Functional Writing test, the same thing is likely to happen; but if the results are good, the impact is likely to remain minimal.

The situation in regard to the CAT is a bit trickier. What the data show is that while the Grade 3 and 5 elementary teachers average 13.5 hours of preparation for this test, eighth grade teachers spend an average of only one hour. Perhaps this is due to the fact that the instruction of a class of students in an intermediate school is the joint responsibility of a group of teachers, none of which feel that they can be held personally accountable for high or low test scores of given students. Or, perhaps the difference is caused by there being 1) much less flexibility in the curriculum at the eighth grade level and 2) there not being a tradition (except, perhaps, in middle schools) of teachers from different subject areas coordinating their efforts. A third possibility is that intermediate school teachers feel that most students have already taken similar tests at the third and fifth grade levels and therefore do not need special preparation, especially in test-taking skills. In any case, there is a striking difference in these data, and future interpretations of county score trends may have to take this factor into account.

EXHIBIT 4

Tests Given in the Secondary Schools*

	No. of Schools**	% of Schools	Adm/ Year	Average Administration Time*** 6	7	8	9	Average Preparation Time 6	7	8	9	Average Time Devoted to Testing**** 6	7	8	9
State-Required Tests															
California Achievement Tests	6	100	1		6.5			-	1.0	-			7.5		
Maryland Functional Mathematics Test	6	100	1	1.5		1.5		0.5	-	0.5		2.0		2.0	
Maryland Functional Reading Test	6	100	1-2	2.0		4.0		3.0	-	2.5		5.0		6.5	
Maryland Functional Writing Test	6	100	1	1.5		1.5		0.0	-	0.0		1.5		1.5	
MCPS-Required Tests															
AAHPERD	6	100	1		2.0			-	0.0	-			2.0		
JOB-O	6	100	1		1.5			-	0.0				1.5		
Writing Proficiency Test	3	50	1			1.5		-	-	1.0				2.5	
Optional Curriculum-Related Tests															
IPR/LA	6	100	1-2	3.0	3.5	3.0		0.5	1.0	0.5	-	3.5	4.5	3.5	
Map and Globe Skills	2	33	1	1.5	1.0	1.0		0.0	0.0	0.0		1.5	1.0	1.0	
Science Criterion-Referenced Test	3	50	1	1.0	1.5	1.5		0.0	0.0	0.0	-	1.0	1.5	1.5	
Optional															
Gates-MacGinite Reading	1	17	1		2.5	2.5		-	0.0	0.0	-		2.5	2.5	
Orleans-Hanna Algebra Prognosis	3	50	1		1.0	1.0	1.0	-	1.0	1.0	0.0		2.0	2.0	1.0
Stanford Diagnostic Reading Test	3	50	1		2.5	2.5	1.0	-	0.0	0.0	0.0		2.5	2.5	1.0
Stanford Achievement Test	1	17	1		8.0	8.0		-	0.0	0.0	-		8.0	8.0	
Other School-Related Tests															
Mathematics	1	17	1			1.0		-	-	0.0	-			1.0	

*Excludes tests for gifted and talented screening and for assessing individual students with special needs.
**For a given test, the number of schools at different grades may vary.
***Rounded to nearest half hour.
****Includes both preparation for testing and actual test administration.

DEA concerns at the intermediate school level are very similar to those voiced for the elementary schools. They, again, relate to questions regarding test quality, using standardized achievement tests for individual student placement purposes, and superimposing additional norm-referenced achievement tests on the already sizable load. In one instance, it was found that an additional eight hours of testing was occurring in both the seventh and eighth grades thanks to each student being given the Stanford Achievement Tests in the fall and the spring. The quality issue is of more widespread importance at the secondary level, as a number of the required tests--specifically the Project Basic Tests, JOB-O, and some of the curriculum-related tests--are of unknown reliability and validity.

A possibility for reducing the burden in the intermediate schools relates to the seventh grade Maryland Functional Reading Test. Although classified as being "state mandated," and treated as such in all Maryland school systems, the present state bylaw does not actually require that the seventh grade MFRT be administered. Given our high passing rate for the ninth grade MFRT, a decision to administer the seventh grade version only to students whose grades and other test scores indicate that they may be "at risk" of failing the ninth grade version might be considered.

In summary, in the secondary schools as in the elementary schools, the actual burden posed by periodic testing amounts to less than 2 percent of the available instructional time. Aside from the state-mandated program, the lion's share of additional testing is done at the option of the local schools. Nonetheless, there are some concerns relating to how some of these tests are used and whether they can be considered "good" measures of the areas they are intended to assess.

OVERALL CONCLUSIONS

The study shows that the amount of periodic testing occurring at both the elementary and secondary levels is not large and in fact consumes less than 2 percent of the available instructional time.

Nonetheless, there are three areas of concern which emerge from these analyses. First, some staff are very concerned over the status of the MCPS-developed curriculum-related tests. It is not the time required for testing which causes these worries as much as what the cumulative time burden in the future might be, given teacher perceptions that they are of uncertain usefulness and quality. We would anticipate that if appropriate in-service training is provided and if the curriculum-related tests can prove their worth, many of the perceived problems with "testing overload" will disappear. These matters should be addressed before use of new tests is expanded systemwide and any mandatory requirements for their use are imposed.

Second, schools continue to misuse achievement test data by using results to place individual students in particular programs and classes. Some schools have also added additional achievement tests to their testing program, whose results cannot be easily interpreted vis-a-vis either the mandated tests or the MCPS Program of Studies. Continued training is needed in the use of norm-referenced achievement tests and the manner in which such tests should be selected.

Finally, questions of reliability and validity can be raised regarding many measures used, especially at the intermediate level. Many are of unknown quality, and their characteristics need to be explored more fully before actions are taken based on their results.

708b/75

APPENDIX A

TECHNICAL ADEQUACY OF TESTS CITED IN THIS STUDY

In this section we discuss briefly the technical adequacy of the tests discussed in this report. The analyses presented are based entirely on the information provided by the test publishers or developers themselves. Because of time limitations, it was not possible for DEA to conduct additional research on the tests' technical properties. The tests will be grouped as they are in the tables.

California Achievement Tests - Most of the CAT subtests have adequate reliability (.80 to .89) for use as group data but probably not for individual decisions. The various subject area totals (combinations of subtests) have good enough reliability (.90+) to be used as part of the information for making decisions about individual students.

A major weakness of the Grade 3 CAT is the ceiling effect. This weakness is especially serious in MCPS because of many high achieving students. On the reading and language subtests in Grade 3, from 30 to 50 percent of MCPS students have scores made inaccurate by the ceiling effect.

Maryland Functional Mathematics Test - There are no technical data available. Some of the items are of questionable quality, because they provide clues that can be used to answer them or other items.

Maryland Functional Reading Test - There are no technical data available. Some of the items are of questionable quality. For example, there are map reading items that use maps of specific parts of the state. These could be biased in favor of students from those places.

Maryland Functional Writing Test - There are no technical data available. It appears that there is a heavy reading component to the Grade 7 level of this test. Thus, it will be hard to determine if a low score at this level is caused by reading or writing problems.

Another potential problem with this test is centered around the fact that the Grade 7 level is multiple choice and on the Grade 9 level the student has to write. The problem arises because the Grade 7 test is used to identify students who need special help before taking the Grade 9 test. If there is little relationship between the skills measured by the two tests this identification could be faulty. Since different skills are being measured there is the possibility of a poor relationship. At this time there are no data available to address this issue.

American Alliance for Health, Physical Education, Recreation and Dance Tests - The AAHPERD tests are designed to measure cardiorespiratory function, body composition, and musculoskeletal function. The actual measures are a distance run (1 to 1½ miles), skinfold fat measures, modified sit-ups, and sit and reach. The test authors present acceptable validity and reliability data for the last three tests. The validity is based on correlations of at least .70 with other measures of the same thing and studies that relate good fitness to these tasks. The test retest reliability coefficients range from .68 to the

high .90's. Unfortunately, no indication of the time between test and retest is presented. Some of the lower reliabilities could have been caused by not enough time (i.e., no rest) or too much time (i.e., time to practice) between the two administrations.

Cognitive Abilities Test - This test has excellent statistical qualities including reliability coefficients that range from .93 to .96. The major problem with it is that it was designed as an ability test but student performance is heavily influenced by what has been learned. Thus, performance is greatly affected by a student's background. This makes interpretation as an ability test highly questionable.

Writing Proficiency Test - An item tryout and a statistical analysis were done when the test was developed. Minor revisions were made based on this. There are no reliability data available.

Instructional Program in Reading/Language Arts - The IPR/LA tests have demonstrated that they will generally produce consistent mastery decisions. The tests also have other good statistical qualities. However, comments from teachers indicate there are problem areas. These include vocabulary in items being too difficult, test format, difficulty in interpreting the results, and short time between testing. While these are not statistical issues, they are probably just as important if the tests are to be accepted and used properly. None of the other tests discussed in the interviews brought forth these kinds of comments.

Map and Globe Skills Test - This test is still in the pilot stage. The first year of the pilot resulted in the need for extensive revision and a second pilot year. None of the data from this second year have been analyzed yet.

Science Criterion-Referenced Tests - There are no data available for the tests that are in the schools. These tests were developed by first having a pilot test. The results from this pilot indicated the need for extensive revision. The tests that are now in the schools are the result of that revision.

School-Initiated Tests - Data were available for several of the tests administered at the option of individual schools. These tests were the Boehm Test of Basic Concepts, Botel Reading Inventory, Clymer-Barrett Prereading Inventory, Gates-MacGinite Reading Test, Metropolitan Readiness Test, Orleans-Hanna Algebra Prognosis Test, Stanford Achievement Tests, and Stanford Early School Achievement Test. In almost all cases the tests have good to excellent reliability. Two exceptions to this should be noted. The publisher of the Boehm did separate reliability studies on students from low, middle, and high socioeconomic-status background. They found generally low reliability for the middle and high groups. This is a fact that should be noted by people considering using this test. The publishers of the Clymer-Barrett noted that the subtests do not have high enough reliability to use for individual students. They recommend using the part scores which combine subtests.

Publishers of several of these tests present validity information that should be considered when deciding if the test should be used. In several cases a claim of content validity is made. For example, the author of the Boehm claims the test contains items related to following directions and understanding oral communication. The content validation of the Botel is

based on the fact that the test is based on a 1968 vocabulary list for words used in basal reading series. For these and the other tests, such claims should be substantiated with respect to their appropriateness for the instructional program for which they will be used.

In other cases the tests have been validated by showing that they do or do not measure the same thing as other tests or indicators. The Metropolitan Readiness Test is to be used to determine if a student is ready for first grade work. The authors used end-of-first-grade test scores as an indication of which students were ready for first grade work. The readiness test showed moderate (.50 to .65) correlations with the end-of-grade test. At this grade level, these correlations are probably pretty good.

A similar analysis was done between Orleans-Hanna scores and end-of-course grades in algebra. These correlations were generally in the 70's which is quite good.

The authors of the Clymer-Barrett wanted to show that their test was a prereading test, not an intelligence test. They correlated it with several intelligence tests and came up with coefficients that ranged from poor (.24) to good (.65). While some of their results were low, the results should be viewed with caution. Low correlations at this age level are common.

Tests used for gifted screening - Reliability data for the Raven, CIRCUS, and Short Form Test of Academic Aptitude are generally good. The only one of these with reliability below .80 is the CIRCUS Think It Through, Level B. No reliability data were available for the Renzulli checklists. A more important issue than reliability for these tests is their validity for determining who should be in the MCPS gifted program. Information would be needed to show that the tests lead to proper placement decisions.

708b/75